สหัส พรหมสิทธิ์

เทคโนโลยีสำหรับค้นหาข้อมูลที่เป็นภาพ

ผลของการวิจัยและพัฒนาที่ก่อให้เกิดกรรมวิธีใหม่ ๆ ในการค้นหาสิ่งที่เป็นภาพหรือสิ่งต่าง ๆ

การค้นหาสิ่งที่เป็นวัตถุธรรมเช่นภาพหรือสรรพสิ่งต่าง ๆ ที่เก็บไว้ในฐานข้อมูลของเน็ตเวอร์กเช่นอินเทอร์เน็ต (เน็ต) นั้นจำเป็นต้องมีวิธีการอันใหม่ที่แตกต่างจาก search engines แบบธรรมดา ซึ่งทำการค้นหาข้อมูลโดยอาศัย**ซอฟต์แวร์บ็อต** (software bots) ที่ท่องไปในเน็ตเพื่อแสวงหาข้อมูลให้ผู้เป็นนายหรือผู้ใช้ข้อมูล อย่างไรก็ตาม บทบาทของบ็อตดังกล่าวอาจต้องสิ้นสุดลงในทันทีที่ต้องไปทำการค้นหาภาพถ่าย แบบเขียน (drawings) ตลอดจนภาพวิดีโอ นับเป็นแสนพันล้านที่มีอยู่ในเน็ต เพราะตัวแทนที่เป็นบ็อตนี้ไม่ได้ถูกสร้างขึ้นมาเพื่อให้ทำการค้นหาสิ่งที่เป็นภาพซึ่งแตกต่างจากสิ่งที่เป็นตัวเลขหรือข้อความหนังสือ (text) ที่ค้นหาได้ง่ายกว่าหลายเท่า จึงมีผู้หาวิธีที่จะหลีกเลี่ยงการสูญเปล่าอันเนื่องมาจาก**ไม่มีการใช้ประโยชน์จากสิ่งที่เป็นภาพ**ที่อยู่ในรูปของออนไลน์ให้ได้เต็มที่ ซึ่งเป็นเนื้อหาสาระสำคัญที่จะพูดในที่นี้ นั่นก็คือผลของการวิจัยและพัฒนาที่ก่อให้เกิดกรรมวิธีใหม่ ๆ ในการค้นหาสิ่งที่เป็นภาพหรือสิ่งต่าง ๆ (เช่น คน สัตว์ ฯลฯ) ที่อยู่ในฐานข้อมูลให้เป็นไปโดยอัตโนมัติ

ค้นหาภาพดิจิตอล

การสร้างอิมเมจอินเด็กซ์ (image index) เท่าที่เป็นอยู่ในขณะนี้ก็คือว่าจะต้องจ้างคนจำนวนหนึ่งมานั่งอยู่หน้าคอมพิวเตอร์ แล้วควานหาภาพต่าง ๆ ในเน็ตทีละภาพเพื่อสร้างแคตาล็อก (cataloging pictures) ขึ้นมา Yahoo! และ search engine รายอื่น ๆ ได้ให้บริการในด้านข้อมูลของภาพ แต่ก็ยังไม่ใช่กิจกรรมขนาดใหญ่และยังต้องใช้แรงคนเข้าช่วย แทนที่จะใช้**ซอฟต์บ็อต** (softbots) สำหรับในการสร้างห้องสมุดของภาพแบบออนไลน์ อันหมายความว่าการใช้บริการค้นหาภาพในฐานข้อมูลที่เป็นอยู่ในขณะนี้ยังคงมีค่าใช้จ่ายสูง เพราะขาดความเป็นอัตโนมัติ อย่างไรก็ตาม ก็ยังมีคอมพิวเตอร์โปรแกรมตามห้องค้นคว้าต่าง ๆ ที่เป็นงานค้นคว้าทดลอง**ในขั้นต้น** เช่น VIR Image Engine ของ Virage Inc. และ Visual SEEK ของมหาวิทยาลัยโคลัมเบีย ที่สามารถค้นหาข้อมูลที่เป็นภาพตามความต้องการของผู้ใช้ ซึ่งต้องระบุรายละเอียดต่าง ๆ เช่น สี รูปร่าง และเนื้อหา เช่น ผิวภายนอก (texture)

การตั้งคำถามเพื่อให้มีการหาภาพที่ต้องการจากฐานข้อมูลในกรณีนี้จึงเป็นเหมือนการต้องสเก็ตช์ภาพอย่างคร่าว ๆ ว่าผู้ใช้ต้องการภาพของอะไร คอมพิวเตอร์โปรแกรมอื่น ๆ เท่าที่มีซึ่งนอกจากที่กล่าวข้างต้นก็จะทำหน้าที่แค่เพียงการเปรียบเทียบเนื้อหาภาพที่เป็นตัวอย่าง (sample picture) กับภาพที่มีอยู่แล้วในฐานข้อมูล อันเป็นโปรแกรมที่อนุโลมใช้ได้ในบางกรณี แต่สิ่งที่เป็นความต้องการอย่างแท้จริงก็คือโปรแกรมที่สามารถวิเคราะห์อิมเมจไฟล์ที่ยังจำแนกว่าเป็นภาพที่เกี่ยวกับอะไร โดยต้องสามารถบอกสิ่งที่อยู่ในภาพอย่างอัตโนมัติโดยไม่ต้องมีใครช่วย

ในขั้นแรกนี้จะขอพูดถึงบางแง่มุมในเรื่องของความสำเร็จในการวิจัยที่เกี่ยวกับ Content Based Image Retrieval (CBIR) เพื่อให้ผู้อ่านเห็นว่าล่าสุดของการค้นคว้าในเรื่องนี้ได้ก้าวไปถึงไหนแล้ว จากนั้นจึงจะกลับมาพูดถึงภาพรวมในเรื่องของ CBIR ว่าแนวทางการค้นคว้าในเรื่องนี้จะต้องอาศัยความรู้ความเข้าใจในด้านใดบ้าง

สำหรับในกรณีที่เป็นแง่มุมหนึ่งของความสำเร็จในเรื่อง CBIR ก็มีว่า เมื่อเร็ว ๆ นี้ ได้มีทีมงานของสองมหาวิทยาลัยในสหรัฐฯ คือ เดวิด ฟอร์สิธ (David Forsyth) แห่งมหาวิทยาลัยแคลิฟอร์เนียที่เบิร์กเลย์ และมาการ์เร็ต เฟล็ค (Margaret Fleck) แห่งมหาวิทยาลัยไอโอวา (ทั้งคู่เป็นรองศาสตราจารย์ด้านคอมพิวเตอร์ศาสตร์) ซึ่งได้พัฒนาโปรแกรมในด้าน CBIR ที่สามารถหาภาพของคนที่อยู่ในสภาพเปลือยเปล่าได้อย่างค่อนข้างแม่นยำ และได้ให้ชื่อโปรแกรมเช่นนี้ว่า Naked People Finder (http.cs.berkeley.edu/~daf/people.html) ที่ได้รับการออกแบบให้สามารถค้นหาไฟล์ของฐานข้อมูลที่เป็นภาพและดึงเอาภาพที่เป็นภาพเปลือยออกมาจากไฟล์ดังกล่าว ศาสตราจารย์ฟอร์สิธได้อธิบายว่า ความเป็นเนื้อเดียวกันของสีของผิวสำหรับภาพเปลือยส่วนใหญ่ที่ได้สแกนออกมานั้นจะเป็นก้าวสำคัญในการทดสอบความแม่นยำของ finder algorithm ส่วนเหตุที่ว่าทำไมจึงต้องเข้าไปเกี่ยวข้องกับภาพเปลือยก็เพราะว่า การพัฒนาโปรแกรมแบบนี้จะนำมาใช้กับเน็ตในอนาคต บวกกับความบังเอิญที่ว่าภายในเน็ตนั้นเป็นแหล่งชุมนุมภาพเปลือย จึงน่าจะเป็นการเริ่มต้นในการพัฒนาโปรแกรมที่เป็น CBIR ที่ชอบด้วยเหตุผลที่สุด นอกจากนี้ถ้าจะมีการค้นหาภาพเช่นของคนในฐานข้อมูล ภาพที่ขาดสิ่งปกปิดจะมีความซับซ้อนน้อยที่สุดในแง่ของการใช้อัลกอริธึมทางคำนวณในการค้นหา

การทำงานของ Naked People Finder จะเริ่มจากการมองหาไฟล์ของภาพในฐานข้อมูลที่ส่วนใหญ่ภายในนั้นจะเก็บข้อมูลที่เรียกว่า skin colored pixels ที่เกี่ยวกับสีสันของผิวเนื้อของมนุษย์ และถ้าไม่นำเอาเรื่องของความดำหรือขาวของสีผิวเข้ามาเกี่ยวข้อง สีผิวของมนุษย์จัดว่ามีความคงที่อย่างมาก อันเป็นอรรถาธิบายของศาสตราจารย์มากาเร็ต เฟล็ค อัลกอริธึมที่เรียกว่า Naked People นี้จะค้นหาภาพของคนโดยมุ่งไปในประเด็นของข้อมูลที่ไม่มีสิ่งที่เรียกว่า strong texture

ตัวอย่างของการทดลองในเรื่องนี้พอจะยกให้เห็นก็คือในกรณีเช่นจาก sample set ที่เป็นภาพเปลือยเป็นจำนวน 565 ภาพ รวมทั้งภาพที่เป็น control image (เช่น ทิวทัศน์ สัตว์ ต่าง ๆ ที่ตั้งทางอุตสาหกรรม คนที่สวมใส่เสื้อผ้า ฯลฯ) นับเป็นจำนวน 4,289 รูป โปรแกรมที่เรียกว่า skin filter จะคัดเลือกภาพเปลือยได้ 448 ภาพและภาพที่เป็น control image เป็นจำนวน 485 รูป ซึ่งแสดงว่า **skin filter** มีความแม่นยำในการคัดเลือกภาพเปลือยของคนคิดเป็นร้อยละ 79 ของทั้งหมด ที่จัดว่าไม่เลวนักเมื่อเทียบกับตัวอย่างของภาพที่แตกต่างกันในแง่ของแบคกราวนด์ texture และสี

ความไม่สมบูรณ์ของโปรแกรมที่เฟล็คและฟอร์สิธได้พัฒนาขึ้นมานั้นได้ทำให้เกิดการผิดพลาด เช่นการเข้าใจผิดว่า control image (ทิวทัศน์ สัตว์ ฯลฯ) เป็นภาพเปลือย ซึ่งมีสูงถึงร้อยละ 11.3 อันทำให้ต้องมีคอมพิวเตอร์โปรแกรมอีกอันมาช่วยจำแนกรูปร่างที่เรียกว่า **geometric filter** ที่ใช้ในการค้นหารายละเอียดของภาพที่แสดงผิวเนื้อของคนโดยทำหน้าที่ต่อเนื่องจากโปรแกรมที่เป็น skin filter เพื่อดูว่าเมื่อเจอภาพที่มีเนื้อหาของผิวเหมือนมนุษย์แล้วก็จะดูต่อไปว่าในเนื้อหาดังกล่าวมีส่วนที่เป็น แขน ขา ฯลฯ หรือไม่

การทำงานของ geometric filter เพื่อหาส่วนต่าง ๆ ที่เหมือนร่างกายของคนนั้นจะอาศัยสมมติฐานที่ว่าร่างกายภายนอกของมนุษย์ประกอบด้วยรูปทรงและส่วนต่าง ๆ ที่มีลักษณะเป็นทรงกระบอก (cylinders) คือถ้าโปรแกรมสามารถค้นหาภาพที่เป็นทรงกระบอกซึ่งสีเหมือนผิวของมนุษย์ มันก็จะหาต่อไปอีกว่าในส่วนที่อยู่ถัดไปนั้นจะมีรูปทรงที่เหมือนกับทรงกระบอกเข้ามาเชื่อมโยงหรือไม่ จากนั้นโปรแกรมตัวนี้ก็จะพยายามผสมผสานสิ่งที่เป็นทรงกระบอกต่าง ๆ เข้าด้วยกัน โดยอาศัยข้อจำกัดอันเป็นรายละเอียดของร่างกายคน ซึ่งก็ได้มีการสร้างโมเดลอันนี้เข้าไปในคอมพิวเตอร์ในรูปของ "body plan" ที่ใช้เป็นเกณฑ์ในการพิจารณาสิ่งที่เหมือนร่างกายคน ผลที่ได้จากการพิเคราะห์นี้ก็คือภาพเปลือยของมนุษย์ที่คอมพิวเตอร์สกัดออกมาจากไฟล์ในฐานข้อมูล คอมพิวเตอร์โปรแกรมที่ทำงานในรูปของ two part system นี้อาศัยหลักของสีของผิวบวกกับรูปทรงทางเรขาคณิตที่เกี่ยวข้องกับร่างกายคนที่เคยจดจำไว้จากที่อื่น

รูปที่ 1 การค้นหาสิ่งที่เป็นรูปทรงของมนุษย์ที่เรียกว่า Content Based Image Retrieval (CBIR) จะอาศัยโปรแกรมที่เรียกว่า skin filter ที่ทำหน้าที่วิเคราะห์รายละเอียดของสีอันเป็นผิวพรรณ จากนั้นก็อาศัยโปรแกรมที่เรียกว่า geometric filter ในการสร้างรูปสามมิติเพื่อดูว่ารูปสำเร็จในขั้นสุดท้ายนั้นเป็นร่างกายของมนุษย์หรือไม่

พัฒนาการในการค้นหาภาพ

เป้าหมายในการค้นหาภาพจากห้องสมุดดิจิตอลของมนุษย์ซึ่งเป็นผู้ใช้นั้นกว้างมาก คือ กว้างทั้งเนื้อหาสาระของภาพ และกว้างทั้งจุดมุ่งหมายในการใช้งาน เช่น ใครสักคนอาจต้องการให้ห้องสมุดที่เก็บงานศิลป์ให้รายละเอียดที่เกี่ยวกับจำนวน shade ของสีของผิวหนังที่ใช้ในภาพเขียนที่เป็นของคน หรืออาจต้องการให้ฐานข้อมูลหาภาพถ่ายดาวเทียมเพื่อนำไปใช้ในการกะประมาณผลเก็บเกี่ยวในปีหน้า ฯลฯ ห้องสมุดที่เก็บภาพดิจิตอลจึงมีความกว้างใหญ่ไพศาลในแง่ของการเป็นที่สะสมภาพนานาชนิดเพื่อสนองความต้องการของผู้ใช้ทุกผู้ทุกนามจนแทบไม่มีขอบเขตในการค้นหาภาพหรือสิ่งที่เป็นสามมิติที่อยู่ในภาพ อันทำให้ต้องมีการวิจัยเพื่อหาระบบวิเคราะห์ภาพที่ทำงานอย่างอัตโนมัติ (fully automated image analysis) ซึ่งเป็นคอมพิวเตอร์โปรแกรมที่จะบอกว่าสิ่งที่อยู่ในภาพคืออะไร หรือมีอะไรบ้าง เช่นอาจเป็นสิงโต หรือดอกไม้ ที่เกิดจากการคัดเลือกภาพเป็นจำนวนมากที่มีลักษณะเหมือนสิงโตหรือเหมือนดอกไม้ทั้งหลายที่มีสีม่วง จนกว่าจะได้ภาพที่มีความชัดเจนในขั้นสุดท้าย (ดังในรูปที่ 2) ที่ผู้ใช้ต้องการ

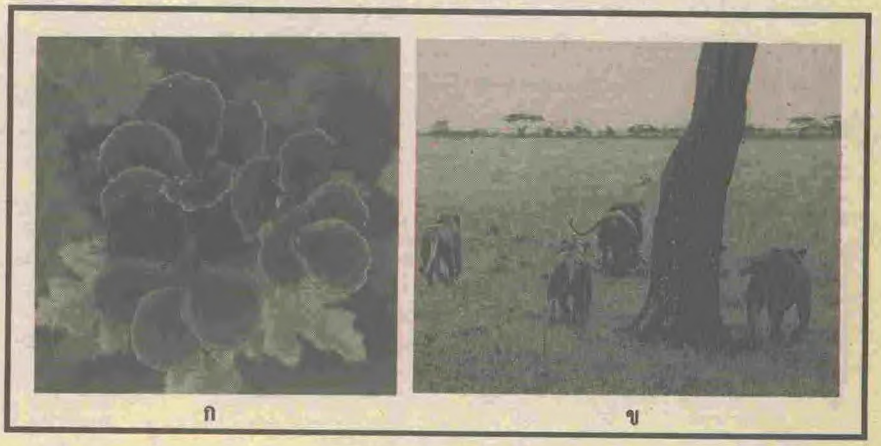
โปรแกรมเช่นที่ว่านี้จะสามารถจำแนกภาพหรือสิ่งที่ต้องการได้อย่างถูกต้อง โดยไม่มีข้อจำกัดในเรื่องของสีสัน ขนาด ลักษณะหรือมุมมอง อย่างไรก็ตาม ในปัจจุบันความเข้าใจที่เกี่ยวกับกรรมวิธีในการหาสาระจากภาพหรือสิ่งสามมิติในภาพยังไปไม่ไกลนัก ซึ่งนักวิจัยตามห้องทดลองต่าง ๆ ของโลกก็ได้ร่วมกันแก้ปัญหาจากในหลาย ๆ ด้าน รวมทั้งการนำความรู้ในด้านคอมพิวเตอร์วิชัน (computer vision) และ automated reasoning ของวิชา A.I. เข้าช่วยในการวิเคราะห์รายละเอียดต่าง ๆ ในภาพ

คอมพิวเตอร์โปรแกรมสามารถวิเคราะห์เนื้อหาของภาพที่จะสกัดออกมาจากไฟล์ในฐานข้อมูลได้หลายวิธี บางโปรแกรมก็จะดูภาพต่าง ๆ เพื่อหาภาพที่เหมือนกับภาพตัวอย่าง ส่วนโปรแกรมอีกพวกหนึ่งก็จะดูลักษณะโดยทั่วไปจากในภาพว่ามีเค้าโครงต่าง ๆ ที่เหมือนกับอะไรในธรรมชาติ (เช่น ทิวทัศน์ ดอกไม้) หรือเหมือนสิ่งประดิษฐ์ของมนุษย์ และปรากฏการณ์ (เช่น รถยนต์ ไฟไหม้ป่า หรือดวงอาทิตย์กำลังจะลับขอบฟ้า) จนถึงโปรแกรมประเภทสุดท้ายที่ทำหน้าที่มองหาสิ่งที่มีเอกลักษณ์ที่ค่อนข้างแน่นอน เช่น ธารน้ำ หรือสิงสาราสัตว์ชนิดใดชนิดหนึ่ง

เทคโนโลยีที่เป็น image retrieval จึงแบ่งออกเป็นสามทางใหญ่ ๆ ด้วยกันคือ

1. หาภาพที่มีลักษณะเหมือนกับตัวอย่าง
2. วิเคราะห์ภาพโดยดูจากองค์ประกอบที่เป็นสีและเนื้อหนังมังสา
3. มองหาความแตกต่างของสิ่งต่าง ๆ เหมือนอย่างที่มนุษย์ทำอยู่

การเปรียบเทียบภาพต่าง ๆ กับภาพที่เป็นตัวอย่างในรูปของ direct matching เป็นกรรมวิธีที่ตรงไปตรงมาแต่มีที่ใช้จำกัด เพราะจะนำมาใช้กับสิ่งที่เป็นสามมิติที่อยู่ในภาพไม่ได้ นอกจากนี้ลักษณะของการปรากฏตัวของสิ่งสามมิติ (เช่น คน สัตว์ ฯลฯ) ในภาพยังมีรายละเอียดอื่น ๆ ให้ต้องคำนึงถึงอีก เช่น ถ้าเป็นสัตว์ก็จะมีลักษณะในการ นั่ง ยืน เดิน หรือวิ่ง ตลอดจนภูมิประเทศอันเป็นแหล่งทำมาหากินของสัตว์แต่ละชนิดเข้ามาให้พิจารณา หรือแม้กระทั่งภาพที่เกิดจากการมองในมุมต่าง ๆ กัน ซึ่งคอมพิวเตอร์โปรแกรมที่ดีจะต้องขบปัญหาให้ตกว่าสิ่งที่เคลื่อนไหวได้ที่อยู่ในภาพจริง ๆ แล้วเป็นอะไรกันแน่ ประโยชน์ของ direct matching จึงตกไป



รูปที่ 2 การใช้ image retrieval program เพื่อหาภาพจากฐานข้อมูลให้เป็นไปอย่างแม่นยำนั้นยังเป็นของยาก เพราะนักวิทยาศาสตร์คอมพิวเตอร์ยังขาดความรู้ความเข้าใจอันลึกซึ้งในด้าน object recognition เช่นการค้นหาพบของดอกไม้สีม่วง (ในภาพ ก) ในฐานข้อมูล ซึ่งผู้ใช้ต้องสอบถาม prototype program เพื่อให้หารูปภาพดังกล่าว ซึ่งก็จะมีภาพของดอกไม้ชนิดต่าง ๆ ที่มีสีม่วงปรากฏออกมาให้เลือกนับเป็นสิบชนิด เพื่อให้ผู้ใช้เลือกสิ่งที่ตรงกับความต้องการจนเหลือเพียงรูปเดียว (ภาพ ก) การค้นหาภาพหรือสิ่งที่ซับซ้อนขึ้นไปอีก เช่นสิงโต คำถามของผู้ใช้จะไปสั่งให้อัลกอริธึมขั้นสูงทำการค้นหาอย่างกว้างไกลในด้านของสีสันและเนื้อหา (texture) จนได้ภาพของสัตว์คล้ายสิงโตดังในภาพ (ข) อัลกอริธึมที่ใช้ในการค้นหาภาพของสิ่งที่มีชีวิตเช่นสัตว์ต่าง ๆ นั้นจะทำงานด้วยความแม่นกว่าอัลกอริธึมโดยทั่วไปที่ไม่ได้นำเอาเรื่องของรูปร่าง (shape) เข้ามาเป็นข้อพิจารณาในการค้นหา หนึ่งในมหาวิทยาลัยที่กำลังมาแรงในเรื่องนี้ที่อยู่ในสหรัฐฯ ก็คือ ม.แห่งแคลิฟอร์เนีย ที่เบิร์กเลย์ ซึ่งเป็นเจ้าของโปรแกรมสองแบบที่ใช้ในการค้นหาภาพข้างต้น (ก) และ (ข)

จากนี้ก็มาถึงการดึงเอาภาพออกมาจากฐานข้อมูล โดยดูองค์ประกอบเนื้อหาในรูปของ "stuff" ที่ดูเหมือนว่าน่าจะมีศักยภาพสูงในการใช้งาน ดังเช่นระบบฐานข้อมูลที่เป็นภาพที่เรียกว่า Query by Image Content (QBIC) ของ IBM ที่เปิดโอกาสให้ผู้ใช้เป็นผู้ระบุรายละเอียดของภาพที่ต้องการจะนำออกมาจากฐานข้อมูลด้วยการบอกสี การกระจายของสี และเนื้อหาอันเป็นรายละเอียด (texture) เช่นลักษณะของสีที่แยกเป็นลายเส้นแบบสลับกันเป็นขาวกับดำ เช่นกรณีของม้าลาย จากนั้นระบบฐานข้อมูลก็จะแสดงลำดับของการคัดเลือกสิ่งที่คิดว่าจะเหมือนกับที่ผู้ใช้ได้กำหนดเข้าไปในคอมพิวเตอร์ เพื่อให้ผู้ใช้เป็นผู้เลือกในขั้นสุดท้าย

อย่างไรก็ตาม การค้นหาภาพโดยอาศัยหลักของ stuff based querries ก็ยังมีข้อจำกัดอยู่อีก เพราะการกำหนดสัดส่วนของสีของสิ่งใดสิ่งหนึ่ง (หรือหลาย ๆ สิ่ง) ที่อยู่ในภาพในขณะตั้งคำถามแก่ฐานข้อมูลยังบอกรายละเอียดไม่หมด เช่น อาจไม่สามารถแยกแยะว่าอันไหนเป็นธงชาติของประเทศใด หากสิ่งที่ปรากฏในภาพเป็นทิวแถวของธงชาติต่าง ๆ ผู้ใช้ที่เป็นมนุษย์มักจะมีความสนใจในสิ่งต่าง ๆ ตามความเป็นจริงที่เป็นรูปธรรม เมื่อเป็นเช่นนี้การเปรียบเทียบเพื่อค้นหาภาพโดยดูจากความละม้ายในแง่ของ "iconic" และ "stuff" จึงยังไม่สามารถที่จะใช้เป็นรากฐานในเรื่องของ content based retrieval ได้ อันทำให้ต้องมีการค้นคว้าวิจัยเพื่อหาระบบค้นหาภาพที่ดีกว่านี้ โดยจะนำผลของการวิจัยในโครงการ Digital Library ของแผนกคอมพิวเตอร์ศาสตร์ของมหาวิทยาลัยแห่งแคลิฟอร์เนียที่เบิร์กเลย์มาพูดต่อไป

งานวิจัยที่เบิร์กเลย์

งานอันเป็นส่วนหนึ่งของโครงการ Digital Library ที่เบิร์กเลย์ก็คือ การวิจัยเพื่อแก้ปัญหาดังกล่าวในข้างต้น งานในช่วงแรก ๆ ที่เกี่ยวกับ image retrieval ที่เบิร์กเลย์เป็นการสร้างระบบเพื่อเก็บภาพกว่า 600,000 ภาพ และ

50,000 ภาพจากจำนวนนี้เป็นภาพที่มาจากที่ต่าง ๆ เช่นภาพพื้นดินและภาพถ่ายทางอากาศ ซึ่งในปัจจุบันเก็บไว้ในรูปของออนไลน์ที่ www ไซท์ของโครงการนี้ หัวใจของเรื่องนี้ก็คือ **อินเทอร์เฟส**ของเบิร์กเลย์ อันเป็นที่รวมคำถามต่าง ๆ เข้าด้วยกัน เพื่อให้ผู้ใช้สามารถค้นหาข้อมูลที่เป็นสิ่งต่าง ๆ (things) และที่เป็นองค์ประกอบของสรรพสิ่ง (stuff) คือ ถ้าหากเป็นการยากที่จะค้นหาสิ่งที่ต้องการก็อาจเลี่ยงได้ด้วยการสั่งให้คอมพิวเตอร์หาภาพที่มีองค์ประกอบ (stuff) ที่สามารถนำมาผสมผสานจนได้สิ่งที่ต้องการในที่สุด โปรแกรมชนิดนี้จึงค้นหาได้ทั้งสองอย่าง (คือ things and stuff) ในเวลาเดียวกัน

ปรัชญาในการค้นหาข้อมูลตามแบบของเบิร์กเลย์จึงเป็นเรื่องของการแสดงออกของสิ่งต่าง ๆ (objects) ที่เกิดจากการเอาสิ่งที่เป็นเนื้อหาหรือองค์ประกอบ (stuff) มารวมกัน อันเป็นการนำไปสู่แนวคิดที่ว่า คำถามในรูปขององค์ประกอบที่มีสาระประโยชน์จะสามารถนำไปใช้สร้างตัวตนอันเป็นรูปร่างของสรรพสิ่งทั้งหลาย อินเทอร์เฟสของผู้ใช้ในกรณีนี้จึงควรมีทางเลือกหลาย ๆ แบบ เพื่อให้ผู้ใช้สามารถเรียกหาข้อมูลที่เป็นสิ่งของ (things) หรือถ้าค้นไม่เจอหรือไม่เป็นที่พอใจก็อาจตั้งคำถามเพื่อถามฐานข้อมูลด้วยวิธีอื่น เช่น การกำหนดส่วนประกอบในรูปของ **spatial assemblies of stuff** ที่สามารถใช้ประโยชน์ได้กว้างกว่าการถามแบบโดยตรง อย่างไรก็ตาม อินเทอร์เฟสที่พัฒนาขึ้นมาในยุคแรก ๆ ที่พูดถึงนี้ยังมีทั้งจุดแข็งและจุดอ่อนในแง่ที่ว่า ภาพที่ได้จากการสอบถามฐานข้อมูลส่วนใหญ่จะตรงกับที่ต้องการ คือเมื่อคิดเป็นร้อยละแล้วจะได้ตัวเลขที่ค่อนข้างสูงในเวลาเดียวกัน ระบบแบบนี้ก็อาจให้ภาพที่ไม่มีความหมายแก่ผู้ถาม หรือไม่ตรงคำถาม ในโลกของการหาข้อมูลที่เรียกว่า information retrieval จึงต้องมีการถัวเฉลี่ยระหว่างความแม่นยำและความสามารถในการระลึกจดจำ (recall) อันหมายถึงว่า เมื่อใดก็ตามที่ต้องการข้อมูลเฉพาะที่จำเป็น ก็จะต้องมีข้อมูลที่หายไปบ้าง และในกรณีที่มีการเรียกข้อมูลทั้งหมดที่ตรงกับที่ต้องการออกมาดู ก็จะมีข้อมูลที่ไม่ต้องการหลงติดมา ซึ่งแสดงว่ากลไกที่เป็น stuff retrieval ที่มีอานุภาพเช่นที่พูดมานี้สามารถนำมาใช้เป็นอุปกรณ์กลั่นกรองเพื่อนำไปสู่การหาภาพที่มีลักษณะเฉพาะแบบที่เป็นตัวเป็นตน (objects) ได้ โดยผู้ใช้ต้องอาศัยดุลยพินิจของตนว่าเมื่อไหร่ควรใช้วิธีใดจึงเหมาะ

การวิจัยและพัฒนาที่เบิร์กเลย์จนถึงขณะนี้จึงมาถึงจุดที่ว่าผู้ใช้สามารถถามข้อมูลที่เป็นภาพได้เพียงไม่กี่อย่าง เช่นภาพของม้า เรือติดเครื่องยนต์ขนาดเล็ก (sailboats) รวมทั้งภาพของคนที่ไม่นุ่งห่ม ผู้ใช้สามารถกำหนดสิ่งที่องค์ประกอบ (stuff) ได้ด้วยการบอกร้อยละของสีชนิดต่าง ๆ ที่คาดว่าจะมีในรูป พร้อมด้วยจำนวนและขนาดของ color blobs ที่จะมี รวมทั้งการกำหนดคุณลักษณะอื่น ๆ ที่แสดงเนื้อหาของภาพ เช่น ภายในภาพมีเส้นแสดงขอบฟ้า (horizon) หรือไม่ สุดท้ายผู้ใช้ยังสามารถที่จะค้นหาข้อมูลที่เรียกว่า metadata อันเป็นข้อมูลที่ติดอยู่กับภาพที่รวมถึงคำบรรยายภาพ ผู้ถ่ายภาพ วัน/เดือน/ปี ที่มีการถ่ายภาพ เป็นต้น

ความยุ่งยากของ object recognition

ถ้าหากมนุษย์ซึ่งเป็นผู้ใช้สามารถให้ความรู้ขั้นพื้นฐานที่ถูกต้องในเรื่องของสรรพสิ่งต่าง ๆ แก่คอมพิวเตอร์ได้อย่างมีประสิทธิภาพ ก็เชื่อได้ว่าคอมพิวเตอร์จะนำความรู้ที่ได้ไป

ทำการวิเคราะห์รายละเอียดของภาพใดภาพหนึ่งว่าประกอบด้วยสิ่งต่าง ๆ อะไรบ้าง การหาเอกลักษณ์ที่แสดงคุณสมบัติเฉพาะของสิ่งมีชีวิต (เช่น คน ม้า สิงโต ฯลฯ) และพวกวัตถุธรรมนั้นเป็นเรื่องของวิชาการที่เรียกว่า object recognition (or) อันเป็นแขนงวิชาที่มีความสำคัญมาก ๆ เรื่องหนึ่งในภาควิชาที่ว่าด้วย computer vision โดยที่ความสำคัญนี้ได้มีมานานหลายสิบปีแล้วในแง่ของการค้นคว้าวิจัย สาเหตุที่ or มีความสำคัญอันค่อนข้างมากก็เพราะประโยชน์ที่ใช้ในทางอุตสาหกรรมและในด้านทหาร เทคนิคเท่าที่ใช้กันอยู่ในขณะนี้ยังมีข้อจำกัดที่ใช้วิเคราะห์ภาพ ซึ่งมีสิ่ง (objects) ที่อยู่ในนั้นได้เพียงไม่กี่อย่างที่รู้รูปร่างหน้าตาอยู่แล้ว โดยบังคับให้มีมุมมองในจำนวนที่จำกัด อันยังไม่ใช่วิธีการที่สัมฤทธิ์ผลที่แม้ว่าจะนำมาใช้กับสรรพสิ่งที่อยู่เดี่ยว ๆ และที่เป็นของธรรมดา เช่น คนหรือสิงโต

สรรพสิ่งที่เป็นของ "ธรรมดา" เหล่านี้มีคุณสมบัติบางอย่างที่บ่งบอกถึงความแปรปรวนได้มาก เช่น ความสูง น้ำหนัก หรือตำหนิต่าง ๆ ที่สามารถหลบหลีกการมองเห็นของ computer vision programs ได้อย่างง่ายดาย จนไม่สามารถแยกแยะข้อแตกต่างเล็ก ๆ น้อย ๆ ในร่างกายของมนุษย์ ในสายตาของคอมพิวเตอร์คนก็คือคน ไม่ว่าบุคคลใดบุคคลหนึ่งจะมีความสูง น้ำหนัก หรือการแต่งกายที่เป็นอย่างใดก็ตาม

ปัญหาที่มีความสำคัญอีกอย่างก็คือการแยกสิ่งต่าง ๆ ที่เป็นเป้าของความสนใจออกจากสิ่งที่เป็นแบคกราวนด์ (backgrounds) ซึ่งถือว่าเป็นเรื่องที่ท้าทายการทำวิจัยเป็นอย่างมาก แม้กระทั่งในการสังเกตสิ่งที่เป็นธรรมชาติแวดล้อมโดยใช้สายตาคน ดังเช่นกรณีของเสือดาวที่แฝงตัวอยู่ในป่าที่เต็มไปด้วยไม้นานาชนิด อันทำให้ต้องมีวิธีจับกลุ่มของภาพที่แลเห็นว่าส่วนใดเป็นลายบนเสือดาว และส่วนใดเป็นใบไม้บนต้นไม้ล้มลุก อันหมายความว่า คอมพิวเตอร์โปรแกรมต้องสามารถจับกลุ่มที่องค์ประกอบที่เรียกว่า stuff เข้าด้วยกันให้เป็นสรรพสิ่ง (things) ในรูปของ perceptual

grouping อันเป็นเรื่องของ human vision ที่อยู่ในวิชาจิตวิทยาในสาขาที่เรียกว่า the Gestalt school of psychology

แนวคิดของ Gestalt ได้ชี้ให้เห็นปัจจัยหลาย ๆ อย่างที่สามารถนำไปใช้หาว่าเมื่อใดที่ส่วนต่าง ๆ ของภาพหนึ่งจะเกิดจากสิ่งใดสิ่งหนึ่งที่อยู่ในทิศทางที่มองออกไป (scene) ความคล้ายคลึงกันในเรื่องของสี หรือผิวพรรณเนื้อหนัง จึงมีความสำคัญมากสำหรับในแง่ของการสังเกตมอง เช่น เผ่าพันธุ์ต่าง ๆ ของมนุษย์ที่สามารถจัดเป็นกลุ่ม โดยดูจากภาพที่เป็นองค์ประกอบซึ่งมีความสม่ำเสมอในเรื่องของสีของผิวอันเข้มคมที่มีความแตกต่างอย่างชัดจากสิ่งแวดล้อม หรืออาจเป็นสีที่เข้ากับธรรมชาติแวดล้อมได้เป็นอย่างดี สำหรับในระดับที่ซับซ้อนขึ้นไปอีกหน่อยในแง่ของการมองเห็นก็คือว่า ย่านต่าง ๆ ในภาพที่เป็นสองมิติที่เป็น symmetric รอบแกนใดแกนหนึ่งนั้นสามารถที่จะรวมกันเป็นกลุ่มของภาพที่เกิดจาก projection ของรูปทรงหรือสรรพสิ่งที่เป็นสามมิติเช่นแจกัน

นักวิจัยในด้าน computer vision ได้ใช้ความพยายามอยู่นานเพื่อหาทางแปลงกฎเกณฑ์ดังกล่าวในข้างต้นให้เป็นอัลกอริธึมสำหรับใช้งาน ดังเช่นงานวิจัยที่เบิร์กเลย์ที่ทำให้เกิดมีระบบที่สามารถแยกภาพใหญ่ออกเป็นภาพเล็ก ๆ ที่เป็นส่วนย่อย ๆ ที่มีความคล้ายกันในเรื่องของสีและผิวเนื้อ ที่แต่ละส่วนดังกล่าวจะมีความกลมกลืนต่อเนื่อง (coherent in color and texture) ในรูปของ "blob" ที่เป็นตัวแทนของคุณสมบัติต่าง ๆ ในเรื่องของตำแหน่ง (เช่นการยืน) รูปร่าง สี และเนื้อหนัง ในกรณีนี้บล็อบ (blobs) จึงหมายถึงสิ่งที่เป็นตัวแทนซึ่งเป็นที่เก็บคุณสมบัติอันเป็นองค์ประกอบที่เป็นพื้นฐานของภาพ (basic compositional features of the image) อันที่จริงแล้วการทำหน้าที่ของ "บล็อบ" ก็อาจเหมือนโปรแกรมเล็ก ๆ หรือส่วนของโปรแกรมที่เก็บคุณสมบัติขั้นพื้นฐานของขอบแต่ละอย่างเอาไว้ เช่น บล็อบของสิงโต บล็อบของดอกไม้แต่ละชนิด บล็อบที่ใช้เป็นหน่วยพื้นฐานของสิ่งที่เรียกว่าเครื่องบิน ฯลฯ ซึ่งผู้ใช้ซอฟต์แวร์แบบนี้สามารถที่จะทำให้แต่ละบล็อบเก็บข้อมูลที่ละเอียดเท่าใดก็ได้ในแง่ของสี เนื้อหนัง รูปร่าง ตลอดจนตำแหน่งการวางตัวของสิ่งต่าง ๆ ทั้งที่เป็นสัตว์ พืช คน และ ฯลฯ

การแทนสิ่งที่เป็น objects ด้วยส่วนย่อย ๆ ที่มีคุณสมบัติจำเพาะเช่นในข้างต้นนี้สามารถที่จะเข้ากันได้อย่างดีกับเทคนิคของ machine learning อันเป็นการใช้สถิติศาสตร์ (statistical learning) เข้าช่วยในการเรียนรู้แทนการใช้ความคิดของคน เพื่อนึกถึงสิ่งที่เป็นจุดเด่นต่าง ๆ ที่มีอยู่ในภาพ (scene) จุดเด่นดังกล่าวมีทั้งสิ่งมีชีวิตและไม่มีชีวิต ที่ทำหน้าที่เป็นตัว "เอก" ของภาพที่เรียกว่า scene กับใช้หลักทางสถิติศาสตร์เข้าช่วยนี้ได้เป็นผลให้มีการพัฒนาซอฟต์แวร์ที่เรียกว่า statistical classifier โดยจะมีตัวอย่างภาพของสิ่งใดสิ่งหนึ่งที่เป็นการเฉพาะ (เช่นภาพของเครื่องบินหรือสิงโต) รวมทั้งการฝึกฝนเพื่อ classifier ดังกล่าวสามารถเข้าใจภาพที่ไม่เคยเห็นมาก่อน ด้วยการจัด "ภาพ" ที่ไม่เคยเห็นมาก่อนให้เป็นส่วนหนึ่งที่เหมือนกับ "ตัวอย่าง" (instances) ที่เคยเก็บไว้ สิ่งที่ห้องทดลองของเบิร์กเลย์ทำได้ในขณะนี้ก็คือว่า classifiers ที่มีอยู่นั้นสามารถจำแนกประเภทของ scenes โดยเพียงแต่ดูข้อมูลในด้านของสีและเนื้อหา และจากข้อมูลที่เหมาะสมในเรื่องของ region shapes ก็จะทำให้ classifiers สามารถแยกกลุ่มของ regions ที่รวมกันเป็น objects รวมทั้งกลุ่มของ regions ที่ไม่ได้เป็น

การที่คอมพิวเตอร์หรือ classifiers เรียนรู้ได้เองนี้จะเป็นเหมือนกรอบทางธรรมชาติที่จะเผชิญกับความแปรปรวนที่เกิดขึ้นในชั้นต่าง ๆ ของ objects ในแง่ของการสามารถปรับตัวเองเพื่อจะได้ไม่ต้องคำนึงถึงการเบี่ยงบ่ายต่าง ๆ ที่ไม่สำคัญ (เช่นท้องฟ้าอาจมีสีฟ้าที่มีความแก่อ่อนของสีในระดับต่าง ๆ หรือลักษณะของก้อนเมฆอาจแตกต่างกันออกไปโดยไม่มีที่สิ้นสุด ฯลฯ)

ทฤษฎีที่ได้จากเบิร์กเลย์และงานวิจัยที่ CMU

ในอีกมุมของการค้นคว้าวิจัยก็ได้มีทีมงานของมหาวิทยาลัยคาร์เนกี เมลลอน (CMU) แห่งสหรัฐฯ ซึ่งนำโดย นายทาเคโอะ คานาเด้ (Takeo Kanade) ซึ่งได้พัฒนาซอฟต์แวร์โมดูลสำหรับเปรียบเทียบใบหน้าคน (face recognition) ที่น่าเป็นตัวอย่างอันดีในแง่ของการนำ learning techniques มาใช้ให้เป็นประโยชน์ในแง่ของการจดจำและเปรียบเทียบใบหน้า โดยได้ฝึกให้นิวรอลเน็ตเวอร์กสามารถจดจำสิ่งที่เป็น ตา จมูก และปากคนให้ได้อย่างแม่นยำ (รวมทั้งสิ่งที่ไม่ใช่ใบหน้าคน) เพื่อประโยชน์ในการเปรียบเทียบในเวลาใช้งาน นอกจากนี้ทีมงานของคานาเด้ยังได้ศึกษาการใช้ประโยชน์จากข้อมูลที่เป็นวิดีโอและเสียง เพื่อวิเคราะห์สิ่งที่อยู่ในภาพวิดีโอและในภาพยนตร์ ในแง่ที่ว่าคอมพิวเตอร์อาจทำหน้าที่ในด้าน speech recognition สำหรับเสียงที่เกิดจากข่าวในเครื่องรับโทรทัศน์ รวมทั้งการจับคู่ระหว่างชื่อของบุคคลสำคัญกับใบหน้าที่เห็นในข้อมูลวิดีโออย่างพร้อมกันไป จากนี้จะขอพูดเรื่อง color และ texture เพื่อเป็นการสรุป

สีและเนื้อหาสามารถเข้ามาช่วยในการหาสิ่งที่เรียกว่า image regions ในแง่ที่ว่าสิ่งต่าง ๆ ที่เป็น objects นั้นจะประกอบด้วยส่วนย่อย ๆ ที่มีรูปทรงที่เป็นแบบสามมิติ รวมทั้งความสัมพันธ์ของรูปทรงดังกล่าวที่เป็นไปอย่างไม่ซับซ้อน นอกจากนี้รูปทรงที่เป็นแบบสามมิติเช่นที่ว่านี้มักจะมีความเรียบง่าย เช่นในรูปของทรงกระบอก (cylinder) อันทำให้คอมพิวเตอร์โปรแกรมสามารถเห็นได้ง่าย ในขณะเดียวกันสำหรับในโลกแห่งความเป็นจริงที่อยู่ภายใต้กฎของธรรมชาติก็จะเป็นว่า จะต้องมีข้อจำกัดในเรื่องของความสัมพันธ์ต่าง ๆ ของรูปทรงทางเรขาคณิตดังกล่าวที่เหมือนกับการเคลื่อนไหวในช่องอันจำกัดของข้อต่อต่าง ๆ ในร่างกายคนและสัตว์ที่เข้ามามีผลในเรื่องของความเป็นไปได้ของความสัมพันธ์ต่าง ๆ ระหว่างแต่ละส่วนของร่างกาย โดยสรุปได้ว่าแต่ละ image region จะมีสมมติฐานต่าง ๆ ที่เป็นของตัวเอง โดยสมมติฐานดังกล่าวจะทำหน้าที่กำหนดกลยุทธ์ในการจับกลุ่มครั้งใหม่เพื่อนำไปสู่การค้นหากลุ่มของ image regions ที่มีขนาดใหญ่ขึ้นและมีความชัดเจนในหน้าที่มากขึ้น ตรงจุดนี้เองที่ statiscal learning theary จะสามารถให้กลไกที่สามารถบอกว่าในกรณีใดกรณีหนึ่งนั้นควรใช้หรือไม่ควรใช้สมมติฐานใดบ้าง จนกระทั่งถึงจุดที่ได้มีการจับกลุ่มกันจนมีขนาดใหญ่พอสมควรด้วยวิธีการดังกล่าวก็จะสามารถชี้ชัดได้ว่าสิ่ง (object) ที่คอมพิวเตอร์แลเห็นนั้นเป็นอะไร แนวทางการค้นคว้าของศาสตราจารย์ฟอร์สิธและมากาเร็ต เฟล็ค ที่กล่าวในตอนแรกก็อาศัยหลักการเช่นนี้ อันเป็นทฤษฎีที่ได้ไปจากห้องทดลองที่เบิร์กเลย์

สรุป

ผลของการค้นคว้าวิจัยเท่าที่พูดในข้างต้นยังขาดความสมบูรณ์ แม้ว่าจะเป็นงานวิจัยที่ทันสมัยที่สุดในแง่ของ Content Based Image Retrieval (CBIR) เพราะในโลกของการค้นหาข้อมูลเช่นการค้นหาภาพนี้จะไม่มีคำว่า perfect search process อันหมายถึงการค้นหาสิ่งที่ผู้ใช้ต้องการโดยไม่มีความบกพร่อง ซึ่งในหมู่ผู้ทำวิจัยในเรื่องนี้มีความเห็นว่าในขณะนี้ยังทำไม่ได้ การค้นหาที่เป็นแบบอัตโนมัติที่ถือว่าดีที่สุดนั้นยังสามารถค้นหาสิ่งที่เป็น indexed material (ที่คิดว่ารัดกุมแล้ว) โดยจะได้สิ่งที่ต้องการไม่เกินร้อยละ 50 และถ้าในกรณีที่ใช้คนเข้าช่วยในการเก็บรวบรวมภาพต่าง ๆ ด้วยมือก็ยังไม่สมบูรณ์อยู่ดี คือไม่สามารถให้ภาพที่ต้องการเกินประมาณร้อยละ 80-90 ท่ามกลางค่าใช้จ่ายอันสูง